

# Generating Logical Specifications from Requirements Models for Deduction-based Formal Verification

Radosław Klimek

AGH University of Science and Technology,  
al. A. Mickiewicza 30, 30-059 Krakow, Poland  
rklimek@agh.edu.pl

**Abstract.** The work concerns automatic generation of logical specifications from requirements models. Logical specifications obtained in such a way can be subjected to formal verification using deductive reasoning. Formal verification concerns correctness of a model behaviour. Reliability of the requirements engineering is essential for all phases of software development processes. Deductive reasoning is an important alternative among other formal methods. However, logical specifications, considered as sets of temporal logic formulas, are difficult to specify manually by inexperienced users and this fact can be regarded as a significant obstacle to practical use of deduction-based verification tools. A method of building requirements models using some UML diagrams, including their logical specifications, is presented step by step. Organizing activity diagrams into predefined workflow patterns enables automated extraction of logical specifications. The crucial aspect of the presented approach is integrating the requirements engineering phase and the automatic generation of logical specifications. A system of the deduction-based verification is proposed. The reasoning process could be based on the semantic tableaux method. A simple yet illustrative example of the requirements elicitation and verification is provided.

**Keywords:** requirements engineering; formal verification; logical specifications; temporal logic; deductive reasoning; semantic tableaux method; use case diagrams; use case scenarios; activity diagrams; workflows patterns;

## 1 Introduction

Software modeling enables better understanding of the domain problem and the system under development. Requirements engineering is an significant part of software modeling. Requirements elicitation should lead into a coherent structure of requirements and have fundamental impact on software quality and costs. Thinking of requirements must precede the analysis, design, and code generation acts. Requirements models are descriptions of delivered services in the context of operational constraints. Identifying software requirements of the system-as-is, gathering requirements and formulation of requirements by users allows defects to be identified earlier in a software life cycle.

UML, i.e. the Unified Modeling Language [6,32], which is ubiquitous in the software industry can be a powerful tool for the requirements engineering process. Use cases are central to UML since they strongly affect other aspects of the modeled system and, after joining the activity diagrams, may constitute a good vehicle to discover and write down requirements. Temporal logic is a well established formalism which allows to describe properties of reactive systems, also visualized in UML. The semantic tableaux method, which is a proof formalization for assessing logical satisfiability, and which might be descriptively called “satisfiability trees”, seems intuitive and may be regarded as goal-based formal reasoning.

Formal methods enable precise formulation of important artifacts arising during software development and help eliminate ambiguity. There are many examples of successful application for formal methods, e.g. [1], and there are challenges for the future [23]. There are two well established approaches to formal reasoning and system verification [13]. The first is based on state exploration (“model checking”) and the second is based on deductive reasoning. However, model checking is an operational rather than analytic approach [12]. Deductive inference enables the analysis of infinite computation sequences. On the other hand, one important problem of the deductive approach is the lack of automatic methods for obtaining logical specifications considered as sets of temporal logic formulas. Even in the case of an average-size system, it consists of many formulas and it is not possible to build a logical specification manually, which can be recognized as a major obstacle to untrained users. Thus, the automation of this process seems particularly important. Moreover, application of the formal approach to the entire requirements engineering phase may increase the maturity of requirements models.

The main issue addressed in the work is a theoretical idea that is verified by finding a workable and comprehensive solution and analyzing a simple yet illustrative and real example. Theoretical problem relates to the generation of logical specifications for requirements models and the application of deductive approach to formal verification of a model behaviour. Logical specifications are always important for formal methods since they constitute proper, formal and representative images/views of the developed model/system. Consequently, they can be used in many phases of the software development cycle, as formal methods offer numerous application possibilities [44]. On the other hand, deductive reasoning plays an important role in the formal approach as a “top-down” and sustainable way of thinking, with reasoning moving from a more general facts to the more specific ones to reach a logically certain conclusion. Let us consider some arguments in favor of a deductive approach.

- The first and the strong argument is the fact that deductive reasoning enables analyzing infinite sequences of computations.
- Another argument is naturalness and common use of deductive reasoning in everyday life. It also dominates in scientific works where obvious deductions are ubiquitous and represent a rational thought sequence that moves linearly from the premises to the conclusion what resembles our normal reasoning.

- A kind of informal argument is an analogy between natural languages and logical approach, that is the strict application of formal grammatical rules, although not necessary, however raises the quality of statements in a natural language, like, by analogy, there is no doubt that the strict application of logical rules for reasoning increases the quality of verification procedures and makes them more reliable.

Logical specifications and deductive reasoning might help to truly understand software models, as they should be understood, what is fundamental in obtaining trustworthy designs as to understand that realizing a good model is more important than producing code [22]. Hence, logic enables evaluation of arguments, i.e. it contains methods and procedures for checking the “reliability” of arguments. Arguments are statements consisting of evidence and a conclusion. Evidence statements are premises while the conclusion must follow from these premises. Thus, the work is also focused on the deduction-based formal verification, i.e. arguing from the general to the particular. It seems that workflow-oriented requirements models are suitable for this kind of verification. Informally speaking, workflows are focused on processes and are not disturbed by data flows.

## 1.1 Motivations and contributions

The motivation for this work is the lack of tools for automatic extraction of logical specifications from software models. Another motivation, which is associated with the previous one, is the lack of tools and practical applications of deductive methods for formal verification of requirements models. However, requirements models built using use cases and activity diagrams seem to be suitable for such an extraction process. All of the above mentioned aspects of the formal approach seem to be an intellectual challenge in software engineering.

The contribution of the work is a method for building formal requirements models, including their logical specification, based on some UML diagrams. A complete deduction-based system which enables the automated and formal verification of requirements models is proposed. The correctness of a model behaviour is considered. Another contribution is a method for automating the generation of logical specifications. The generation algorithm for some workflow patterns is presented. Although the work is not based on any particular method of reasoning, i.e. generated logical specifications can be used for many purposes and reasoning engines, the semantic tableaux method for temporal logic is suggested. The proposed generating method is characterized by the following advantages: introducing workflow patterns as logical primitives to requirements engineering and logical modeling, scaling up to real-world problems, i.e. migration from small models to real problems in the sense that they are having more and more predefined patterns and more nesting expressions, and logical patterns once defined, e.g. by a logician or a person with good skills in logic, and widely used, e.g. by analysts and engineers with less skills in logic. All these factors are discussed in the work and summarized in the last Section.

## 1.2 Related works

There is a large volume of published works describing the role and importance of requirements engineering, as well as some conferences and journals are dedicated to this subject. There is an unambiguous relationship between the quality of the requirements engineering phase and the quality of developed system. Some fundamental works on requirements engineering are published, c.f. works by Sommerville [40], van Lamsweerde [30], or by Pohl [33] which are comprehensive studies of many fundamentals of this area. Work by Chakraborty et al. [9] discusses some social processes associated with requirements engineering. Work edited by Yu et al. [45] discusses some aspect of social modeling for requirements engineering including modeling framework, applications in security/privacy, incorporating and evaluating social modeling, etc. In work by Winkler and Pilgrim [41] the problem of traceability in the requirements engineering context is discussed. Traceability is understood as the ability to follow the life of software artifacts. Work is a review of research and practice identifying commonalities and differences in these areas. Work by Cao and Ramesh [8] discusses requirements engineering in agile development, and some real cases are considered. This approach is suitable for rapidly changing business environment and differs from traditional approach, thus, it should be used when developing unambiguous and complete requirement specifications is impossible.

There are many approaches for building and analysis requirements models. In the work by Rauf et al. [34], a method for extracting logical structures from text documents is presented, however, this solid work concerns rather recognition instances of use cases, business rules, etc., and discovered relations are not understood in a formal/logical way, i.e. in terms of logical specifications. Work by Kazhamiakin et al. [27] discusses a method based on formal verification of requirements, as well as a case study on web services is discussed, however, linear time logic and model checking is used. Blanc et al. [5] propose a kind of formalism as sequences of elementary constructions that originate from different models, and are considered uniformly in the work, then the logic-based analysis of inconsistency using the Prolog engine is performed. The work constitute an interesting logic-based approach and distinctive dissimilarity to other works/approaches. Work by Nikora and Balcon [31] provides a method for identifying and discovering temporal properties contained in a natural language as requirements. These requirements are specified as temporal logic patterns. Some machine learning techniques are used. Properties can be converted to finite state automata and analyzed using model checking techniques. Work by Smith and Havelund [39] is another work providing tools, that graphically cover formal requirements, and enabling verification using the model checking techniques.

Many works concern formalization, and verification, of the UML diagrams that are used in the work. Hurlbut [24] provides a very detailed survey of selected issues concerning formalization of use cases. The informal character of scenario documentation implies several difficulties in reasoning about the system behavior and validating the consistency between the diagrams and scenario descriptions. Work by Barrett et al. [3] proposes formal definition of syntax and semantics

of use cases to enable modeling of use cases, detecting their inconsistencies and conflicts. Some illustrative examples are presented. Zhao and Duan [46] shows formal analysis of use cases; however, the Petri Nets formalism is used. The solid work by Eshuis and Wieringa [17] addresses the issues of activity diagram workflows but the goal is to translate diagrams into a format that allows model checking. Then, propositional requirements are checked against the input model. Some examples are provided. In the work by Cabral and Sampaio [7], a method for automatic generation of use cases and a proposed subset of natural languages, however, for the algebraic specification is introduced. Work by Rossi et al. [35] provides interesting and comprehensive formalization of state machines using temporal logic, however, only state diagrams are discussed which are not considered in the work. There is a variety of formalisms used in these area, however, they are not discussed widely in the work.

In the comprehensive and general work by Shankar [38] automated deduction for verification using symbolic logical reasoning is widely discussed. There are considered satisfiability procedures, automated proof search, and variety of application in the case of propositional and fragments of first-order logic. However, even though the work contains a survey of symbolic reasoning, modal and temporal logics are not considered widely. As it is already cited in work [35], the statement by Chomicki and Saake taken from [11], “Logic has simple, unambiguous syntax and semantics. It is thus ideally suited to the task of specifying information systems”. Logic offers also many possibilities of applications, i.e. specification, verification, synthesis, and programming [18]. “Logic is the glue that binds together methods of reasoning, in all domains”, thus “we need a style of logic that can be used as a tool in every-day work” [20]. Moreover, “automated deduction tools can be used in a variety of ways in formal verification in applications ranging from modeling requirements and capturing program semantics to generating test cases” [38]. In work [13] some examples of existing theorem provers, i.e. when both the system and desired properties are expressed as logical formulas, with a different degree of automation, are briefly discussed. In the end of Section 4, some works discussing experimental results with automated theorem provers using temporal logic are presented.

Summing up, there are many works in the domain of requirements engineering, and works in the area of the formal approach for the UML-based requirements engineering but there is a lack of works for extracting logical specifications from requirements models and deduction-based formal verification with temporal logic as well as the semantic tableaux method for UML-based requirements models. This work is an extended version of the work [29]. The major differences are that some new formal notions are introduced, the generation algorithm is improved and expressed in a more formal way, as well as more widely discussed, the main example of the work is extended both in the case of modeling and reasoning.

### 1.3 Organization

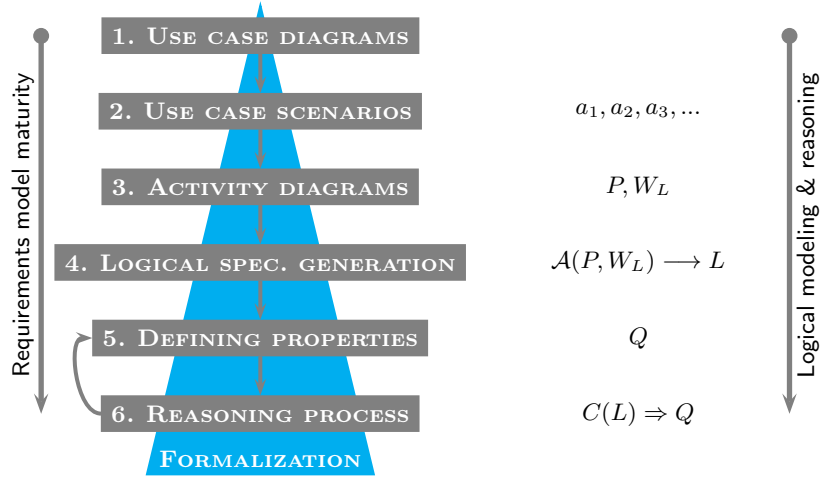
The rest of the work is organized as follows. The procedure and guidelines for the construction of formal requirements models are introduced in Section 2. The presented procedure contains both manually and automatically performed steps. The problem of identifying atomic activities using use case diagrams and scenario is discussed in Section 3. The appropriate example is shown. Logical background which are temporal logic and logical inference using the semantic tableaux method are discussed in Section 4. Temporal logic is an established standard for the specification and verification of reactive systems and the semantic tableaux method is a natural and valuable method of inference. The deduction system and its architecture is also proposed. The system enables formal verification of requirements models. Workflow patterns for activity diagrams are introduced in Section 5. They are treated as (logical) primitives which allow to automate the entire process of generating logical specifications. Launched real example is continued. The algorithm for generation logical specifications is proposed in Section 6. Introduced workflow are predefined in terms of temporal logic formulas. Some properties of the proposed generation algorithm is discussed in the end of the Section. The general example for the requirements model considered in previous Sections is continued in Section 7 by the generating logical specifications for the model and formal analysis of its behavioral correctness. The work is summarized and further research are discussed in Section 8.

## 2 Towards a methodology

The outline of the procedure and guidelines used for the construction of a requirements model, as it is understood in the work, is briefly discussed below. It constitutes a kind of methodology and its subsequent steps are presented in Fig. 1. The entire procedure can be collected in the following items:

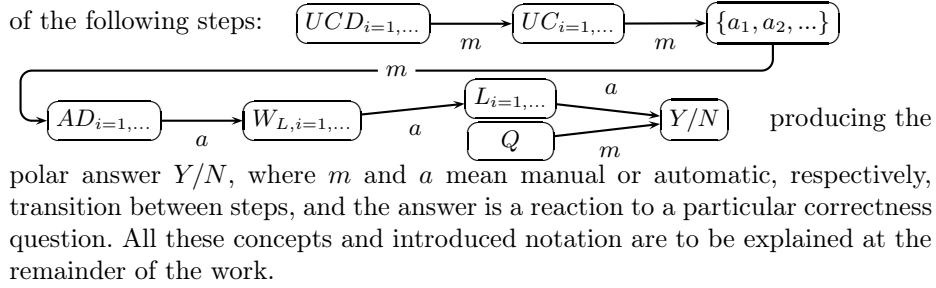
1. use case diagrams – use case modeling to understand functions and goals of a system;
2. use case scenarios – identifying and extracting atomic activities;
3. activity diagrams – modeling workflows using predefined patterns;
4. automatic generation of logical specifications from requirements models;
5. manual definition of the desired model properties;
6. formal verification of a desired property using the semantic tableaux method.

All steps are shown on the left side of Fig. 1. The first three steps involve the requirements modeling phase but the last three steps involve generation of logical specification and analysis of requirements model properties. The loop between the last two steps refers to a process of both identifying and verifying more and more new properties of the examined model. Some symbols and notation resulting from the introduced formalization are on the right side of Fig. 1 and they are discussed in further sections of the work. Generally, it leads, step by step, from an abstract view of a system to more and more detailed and reliable and, finally, verified requirements models.



**Fig. 1.** Software requirements modeling and deduction-based verification

Let us summarize the entire method proposed in the work through a sequence of the following steps:

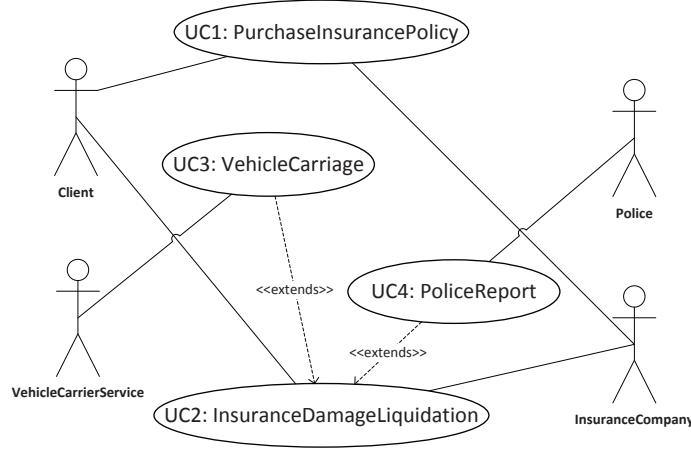


producing the polar answer  $Y/N$ , where  $m$  and  $a$  mean manual or automatic, respectively, transition between steps, and the answer is a reaction to a particular correctness question. All these concepts and introduced notation are to be explained at the remainder of the work.

### 3 Use cases and identification of activities

Defining use cases and scenarios is important not only to understand the functionalities of a system but also to identify elementary activities. The activities play an important role when building logical specifications, i.e. the logical specification is modeled over atomic activities. The *use case diagram* consists of actors and use cases. *Actors* are objects which interact with a system and create the system's environment, thus providing interaction with the system. *Use cases* are services and functionalities which are used by actors. The use case diagrams are a rather descriptive technique and do not refer to any details of the system implementation [37].

Let us present it more formally. In the initial phase of a system modeling, use case diagrams  $UCD_1, \dots, UCD_n$  are built. Every  $UCD_i$  diagram contains some use cases  $UC_1, \dots, UC_m$  which describe the desired functionality of a system. A typical and sample use case diagram is shown in Fig. 2. It consists of four actors



**Fig. 2.** A sample use case diagram  $UCD$  “CarInsuranceLiquidatingDamages”

and four use cases,  $UC_1$ ,  $UC_2$ ,  $UC_3$  and  $UC_4$ , modeling a system of car insurance and damages liquidation. The diagram seems to be intuitive and is not discussed in detail.

Use cases are commonly used for capturing requirements through *scenarios* which are brief narratives that describe the expected use of a system. A scenario is a possible sequence of steps which enables the achievement of a particular goal resulting from the functionality of a use case. Every use case  $UC_i$  has its own scenario. From the point of view of the approach presented here, scenarios play an additional important role, which is identification of atomic activities used to build individual scenario steps. An *activity* is the smallest unit of computation. Thus, every scenario contains some activities  $a_1, \dots, a_n$ . The set of *atomic activities*  $AA$  contains all activities identified and defined for all scenarios. The most valuable situation is when the entire use case scenario involves identified activities and the narrative does not dominate and is limited to model behavior, which is later formally shown in activity diagrams.

Sample scenarios for use cases  $UC_2$  and  $UC_3$ , i.e. “InsuranceDamageLiquidation” and “VehicleCarriage”, are shown in Fig. 3. They contain some atomic activities which are identified when preparing scenarios. The alternative and extension points are defined. The “Application” activity represents the act of applying for an initiation of the compensation procedure. “DamageVindication” represents the registration process in the insurer system, verification of insurance, and the start, if justified, of the process of recovery damages. If all documentation is gathered (“SupplyDocumentaryEvidence”), it can start repairing (“InitRepair”). If necessary then the use case is extended by the “VehicleCarriage” use case. After the case registration and checking the insurance scope at the “VehicleCarrierService” actor, the car is transported. Let us note that the “InitRepair” activity is common to  $UC_2$  and  $UC_3$ , which shows a joint point and



<b>UC2: InsuranceDamageLiquidation</b>		
<b>Scenario:</b>		
1.	Client's "Application"	
2.	"DamageVindication"	
3.	"SupplyDocumentaryEvidence"	
4.	"InitRepair"	
5.	"RentVehicle"	
6.	"MechanicalRepairs"	
7.	"BodyRepairs"	
8.	"TestDrive"	
9.	InsuranceCompany's "FinalReport"	
<b>Alternatives:</b>		
If vindication is already processed then		
"WarningDoubleVindication" and		
"FinalReport"		
<b>Extensions:</b>		
1.	If there exists the police report then	
	include it when	
	"SupplyDocumentaryEvidence"	
2.	If necessary then "VehicleCarriage"	
<b>UC3: VehicleCarriage</b>		
<b>Scenario:</b>		
1.	"CallAssistance"	
2.	"TransportVehicle"	
3.	"InitRepair"	

**Fig. 3.** Scenarios for use cases  $UC_2$  "InsuranceDamageLiquidation" (left) and extended  $UC_3$  "VehicleCarriage" (right)

does not mean that the activity is executed more than once. While the car repair process is carried out ("MechanicalRepairs" and "BodyRepairs"), the client can hire a replacement vehicle ("RentVehicle"). At the end of the scenario, it is always generated a report ("FinalReport"). The level of formalization presented here, i.e. when discussing use cases and their scenarios, is intentionally not very high. This assumption seems realistic since this is an initial phase of requirements modeling. Dynamic aspects of activities are to be modeled strictly when developing activity diagrams, c.f. Section 5.

## 4 Logical background

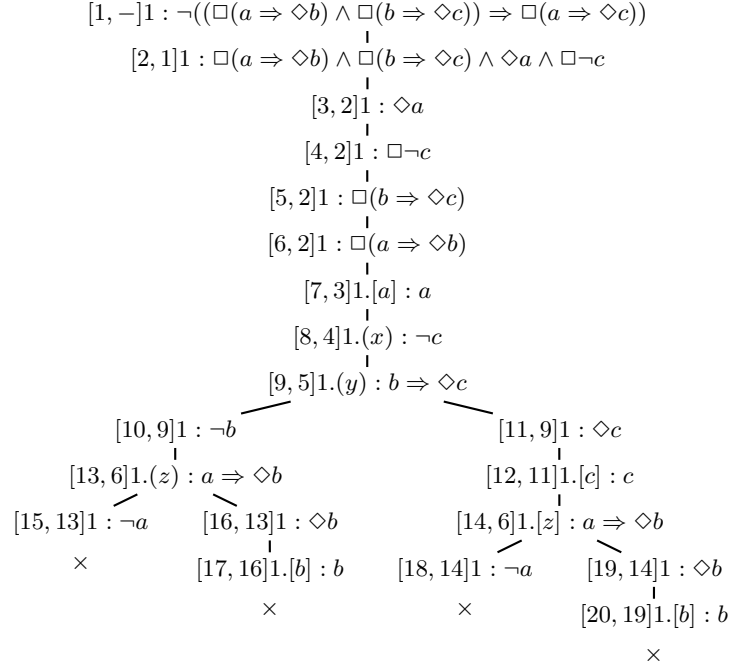
*Temporal logic* TL introduces symbolism for reasoning about truth and falsity of formulas throughout the flow of time, taking the changes of their valuations into consideration. Two basic and unary operators are  $\Diamond$  for "sometime (or eventually) in the future" and  $\Box$  for "always in the future"; these are dual operators. Temporal logic exists in many varieties; however, these considerations are limited to the *linear-time temporal logic* or *linear temporal logic* LTL. Linear temporal logic refers to infinite sequences of computations and attention is focused on the *propositional linear time logic* PLTL. These sequences refer to the Kripke structure which defines the semantics of TL, i.e. a syntactically correct formula can be satisfied by an infinite sequence of truth evaluations over a set of *atomic propositions* AP. It should be pointed out that atomic propositions are identical to atomic activities defined in Section 3, i.e.  $AA = AP$ . The basic issues related to temporal logics and their syntax and semantics are discussed in many works, e.g. [16,43].

The properties of the time structure are fundamental to a logic. Of particular significance is the *smallest*, or *minimal*, *temporal logic*, e.g. [10,4], also known as temporal logic of class K. The minimal temporal logic is an extension of the classical propositional calculus of the axiom  $\Box(\Phi \Rightarrow \Psi) \Rightarrow (\Box\Phi \Rightarrow \Box\Psi)$  and the inference rule  $\vdash \Phi \Rightarrow \vdash \Box\Phi$ . The essence of the logic is the fact that there are no specific assumptions for the properties of the time structure. The following formulas may be considered as typical examples:  $action \Rightarrow \Diamond reaction$ ,  $\Box(send \Rightarrow \Diamond receive)$ ,  $\Diamond alive$ ,  $\Box\neg(badevent)$ , etc. The considerations of the work are limited to this logic since it allows to define many system properties (safety, liveness); it is also easier to build a deduction engine, or use existing verified provers, and to quickly verify the approach proposed in the work.

Although the work is not based on any particular method of reasoning, the method of semantic tableaux is presented in a more detailed way. *Semantic tableaux* is a decision-making procedure for checking satisfiability of a formula. The method is well known in classical propositional logic but it can also be applied in modal and temporal logics [14], and for the propositional linear-time logic first presented in [42]. The method is based on formula decompositions. In the semantic tableaux method, at the end of the decomposition procedure, all branches of the received tree are searched for contradictions. When all branches of a tree have contradictions, it means that the inference tree is *closed*. If a negation of the initial formula is placed in the root, this leads to the statement that the initial formula is true. This method has some advantages over the traditional axiomatic approach. In the classical reasoning approach, starting from axioms, longer and more complicated formulas are generated and derived. Formulas become longer and longer step by step, and only one of them will lead to the verified formula. The method of semantic tableaux is characterized by a reverse strategy. The method provides, through so-called *open* branches of the semantic tree, information about the source of an error, if one is found, which is another and very important advantage of the method. Summing up, the tableaux are global, goal-oriented and “backward”, while resolution is local and “forward”.

A simple yet illustrative example of an inference tree is shown in the left side of Fig. 4. The relatively short formula gives a small inference tree, but shows how the method works. The label  $[i, j]$  means that it is the  $i$ -th formula, i.e. the  $i$ -th decomposition step, received from the decomposition transformation of a formula stored in the  $j$ -th node. The label “1 :” represents the initial world in which a formula is true. The label “1.( $x$ )”, where  $x$  is a free variable, represents all possible worlds that are consequences of world 1. On the other hand, the label “1.[ $p$ ]”, where  $p$  is an atomic formula, represents one of the possible worlds, i.e. a successor of world 1, where formula  $p$  is true. The decomposition procedure adopted and presented here refers to the first-order predicate calculus and can be found, for example, in the work [21]. All branches of the analyzed trees are closed ( $\times$ ). There is no valuation that satisfies the root formula. This, consequently, means that the formula before the negation is always satisfied.

An outline architecture of the proposed deduction-based verification system is presented in Fig. 5. A similar system is proposed in work [28]. The system

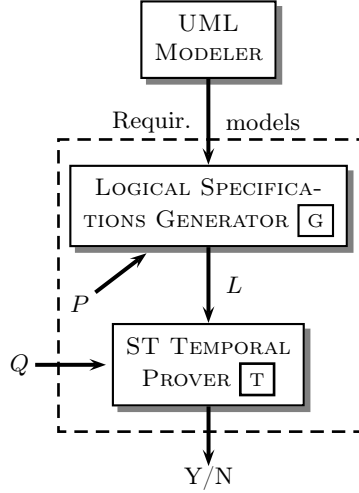


**Fig. 4.** A sample inference tree

works automatically and consists of some important elements. The  $\boxed{\text{G}}$  component generates logical specifications which are sets of a usually large number of temporal logic formulas (of class K). Formula generation is performed automatically from workflow models, which are constructed from predefined patterns for activity diagrams. The extraction process is discussed in Section 6. The whole specification  $L$  can be treated as a conjunction of formulas  $f_1 \wedge \dots \wedge f_n = C(L)$ , where every  $f_i$  is a formula generated during the extraction process. The  $Q$  formula is a desired property for a requirements model. Both the system specification and the examined properties are input to the  $\boxed{\text{T}}$  component, i.e. *Semantic Tableaux Temporal Prover*, or shortly *ST Temporal Prover*, which enables the automated reasoning in temporal logic using semantic tableaux. The input for this component is the formula  $C(L) \Rightarrow Q$ , or, more precisely:

$$f_1 \wedge \dots \wedge f_n \Rightarrow Q \quad (1)$$

Due to the fact that the semantic tableaux method is an indirect proof, i.e. *reductio ad absurdum*, after the negation of Formula 1, it is placed at the root of the inference tree and decomposed using well-defined rules of the semantic tableaux method. If the inference tree is closed, it means that the initial Formula 1 is true. The output of the  $\boxed{\text{T}}$  component, and therefore also the output of the entire deductive system, is the answer Yes/No. This output also realizes the final step



**Fig. 5.** A deduction-based verification system

of the procedure shown in Fig. 1. However, the verification procedure can be performed for the further properties, c.f. the loop in Fig. 1.

The verification procedure which results from the deduction system in Fig. 5 can be summarized as follows:

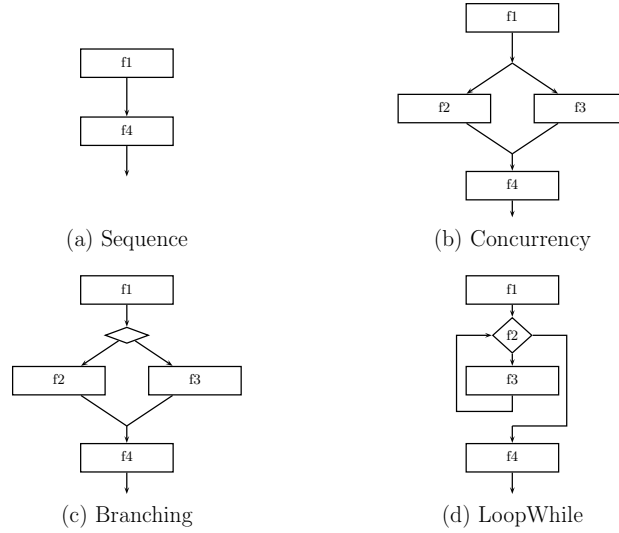
1. automatic generation of system specifications (the  $\boxed{G}$  component);
2. introduction of the property  $Q$  of the system;
3. automatic inference using semantic tableaux (the  $\boxed{T}$  component) for the whole complex formula, c.f. Formula 1.

Steps 1 to 3, in whole or individually, may be processed many times, whenever the specification of the UML model is changed (step 1) or if there is a need for a new inference due to the revised system specification (steps 2 or 3).

The prover is an important component of the architecture for the deduction-based system shown in Fig. 5. It enables automate the inferencing process and formal verification of developed models. Reasoning engines are more available, especially in recent years when a number of provers for modal logics become accessible, c.f. Schmidt [36]. Work by Goré et al. [19] provides experimental results for some existing provers, which are based on different methods of reasoning. Another work by Hustadt and Schmidt [25] provides also an experimental performance analysis of some other theorem provers based on different method of reasoning, and some randomly generated formulas are analysed. Work by Islam et al. [26] provides a brief overview of some existing tableau theorem provers. In work by Dixon et al. [15] an automated tableau prover generator is used and some implementation and experimental results are discussed. Summing up, selection of an appropriate existing prover, or building one's own, constitutes a separate task that exceeds the size and main objectives of the work, c.f. also the concluding remarks in the last Section 8.

## 5 Workflow patterns and modeling activities

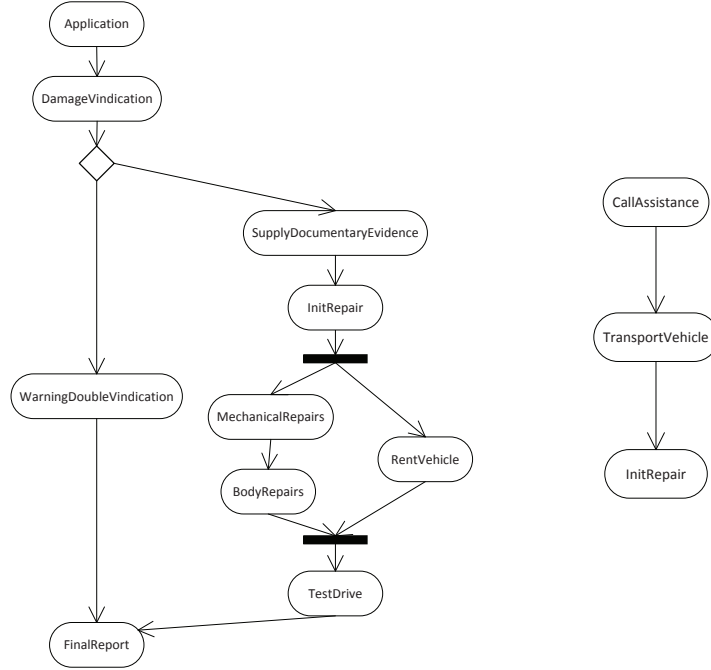
Activity diagrams constitute a closure of the development phase for requirements models, by introducing dynamic aspects for models. This aspect is subjected to the correctness analysis for safety and liveness properties. The *activity diagram* enables modeling of workflow activities. It constitutes a graphical representation of workflow showing the flow of control from one activity to another. It supports choice, concurrency and iteration. The *swimlane* is useful for partitioning the activity diagram and enables grouping of activities in a single thread. The important goal of activity diagrams is to show how an activity depends on others [32].



**Fig. 6.** Workflow patterns for activities

From the viewpoint of the approach presented in the work, it is important to introduce a number of predefined workflow patterns for activities that provide all workflows in a structural form. A *pattern* is a generic description of the structure of some computations. Nesting of patterns is permitted. The following workflow patterns are predefined: *sequence*, *concurrent fork/join*, *branching* and *loop while* for iteration as they are shown in Fig. 6. It is assumed that only predefined patterns can be used for modeling of activity behavior. Such structuring is not a limitation when modeling arbitrarily complex sets of activities.

For every use case  $UC_i$  and its scenario, a activity diagram  $AD_i$  is developed/modeled. The activity diagram workflow is modeled only using atomic activities which are identified when building a use case scenario. Furthermore, workflows are composed only using the predefined design patterns shown in



**Fig. 7.** Sample activity diagrams  $AD_2$  (left) and  $AD_3$  (right) for use cases  $UC_2$  “InsuranceDamageLiquidation” and  $UC_3$  “VehicleCarriage”

Fig. 6. Sample activity diagrams  $AD_2$  and  $AD_3$  are shown in Fig. 7. They model behavior of the  $UC_2$  and  $UC_3$  use cases shown in Fig. 2, using activities from the scenarios in Fig. 3. After the start of the vindication process, i.e. “DamageVindication”, it is checked whether it is already being processed. If yes, the decision to register this fact is made, as it is likely another attempt at vindication of the same event, c.f. “WarningDoubleVindication”. The scenario analysis and the nature of other activities, i.e. “MechanicalRepairs”, “BodyRepairs” and “RentVehicle”, leads to the conclusion that they can and should be performed concurrently.

This step completes the phase of modeling requirements, c.f. Fig. 1. The next steps involve the widely understood formal analysis of obtained requirements models.

## 6 Generating logical specifications

The phase of modeling requirements is complete when all activity diagrams for all scenarios are built, c.f. Fig. 1 and Section 5. Then, the phase of generating logical specifications and formal analysis of the desired properties begins. The logical specification generation process must be performed in an automatic way. Such logical specifications usually consist of a large number of temporal logic formulas

and their manual development is practically impossible since this process can be monotonous, error-prone and the creation of such logical specifications is difficult for inexperienced analysts. On the other hand, the verified properties of the system constitute usually easier formulas, not to mention the fact that they are rather individual temporal logic formulas.

The proposed algorithm for automatic extraction of logical specifications is based on the assumption that all workflows for activity diagrams are built using only well-known workflow patterns, c.f. Fig. 6. The process of building a logical specification can be presented in the following steps:

1. analysis of activity diagrams to extract all predefined workflow patterns,
2. translation of the extracted patterns to a logical expression  $W_L$ ,
3. generating a logical specification  $L$  from logical expressions, i.e. receiving a set of temporal logic formulas.

Predefined workflow patterns constitute a kind of primitives which are defined using temporal logic formulas. Therefore, an *elementary set*  $pat()$  of formulas over atomic formulas  $a_i$ , where  $i > 0$ , which is also denoted  $pat(a_i)$ , is a set of temporal logic formulas  $f_1, \dots, f_m$  such that all formulas are syntactically correct (and restricted to the logic K). For example, an elementary set  $pat(a, b, c, d) = \{a \Rightarrow \Diamond b, b \Rightarrow \Diamond(c \vee d), \Box \neg((a \vee b) \wedge \neg c)\}$  is a three-element set of PLTL formulas, created over four atomic formulas. Let  $\Sigma$  be a set of *predefined design patterns*, i.e.  $\Sigma = \{Sequence, Concurrency, Branching, LoopWhile\}$ . The proposed temporal logic formulas should describe both safety and liveness properties of each pattern. Let us introduce some aliases: *Seq* as *Sequence*, *Concur* as *Concurrency*, *Branch* as *Branching* and *Loop* as *LoopWhile*.

Every activity workflow is designed using only predefined design patterns. Every design pattern has a predefined and countable set of linear temporal logic formulas. The workflow model can be quite complex and it may contain nesting patterns. Let us define a logical expression, which is similar to well known regular expressions, to represent any potentially complex structure of the activity workflow but also to have a literal representation for these workflows. The *logical expression*  $W_L$  is a structure created using the following rules:

- every elementary set  $pat(a_i)$ , where  $i > 0$  and every  $a_i$  is an atomic formula, is a logical expression,
- every  $wrf(A_i)$ , where  $i > 0$  and every  $A_i$  is either
  - an atomic formula, or
  - a logical expression  $pat()$ ,
is also a logical expression.

Examples of logical expressions are given in Section 7.

Some restrictions on set of atomic formulas  $a_1, \dots, a_n$  of the logical expression  $pat()$ , due to the number of arguments and their partial order, are introduced.  $a_1$  is always the first argument called an *entry argument/formula*, and closely there is one entry argument.  $a_4$  is always the last argument called an *exit argument/formula*, and closely there is one exit argument. The subset of arguments

between, informally speaking, the first and the last argument are *ordinary arguments* and this subset may be empty. This implies that there are at least two arguments, but not more than four arguments. From this it also follows that there is also only one entry, or exit, respectively, point for a pattern, c.f. Fig 8. The further discussion on these limitations, especially number of arguments  $n = 4$ , is provided in the end of Section 6. Let us also note that entry and exit formulas enable representing the pattern as a whole, i.e. without analysis of its internal behavior if it is necessary.

```

/* ver. 6.01.2014 */
Sequence(f1,f4):
f1 => <>f4 / ~f1 => ~<>f4 / []~(f1 & f4)
Concurrency(f1,f2,f3,f4):
f1 => <>f2 & <>f3 / ~f1 => ~(<>f2 & <>f3)
f2 & f3 => <>f4 / ~(f2 & f3) => ~<>f4
[]~(f1 & (f2 | f3)) / []~((f2 | f3) & f4) / []~(f1 & f4)
Branching(f1,f2,f3,f4):
f1 => (<>f2 & ~<>f3) | (~<>f2 & <>f3)
~f1 => ~((<>f2 & ~<>f3) | (~<>f2 & <>f3))
f2 | f3 => <>f4 / ~(f2 | f3) => ~<>f4
[]~(f1 & f4) / []~(f2 & f3)
[]~(f1 & (f2 | f3)) / []~((f2 | f3) & f4)
LoopWhile(f1,f2,f3,f4):
f1 => <>f2 / ~f1 => ~<>f2
f2 & c(f2) => <>f3 & ~<>f4
~(f2 & c(f2)) => ~(<>f3 & ~<>f4)
f2 & ~c(f2) => ~<>f3 & <>f4
~(f2 & ~c(f2)) => ~(~<>f3 & <>f4)
f3 => <>f2 / ~f3 => ~<>f2
[]~(f1 & f2) / []~(f1 & f3) / []~(f1 & f4)
[]~(f2 & f3) / []~(f2 & f4) / []~(f3 & f4)

```

**Fig. 8.** A predefined set of patterns  $P$  and their temporal properties

The last step is to define a logical specification which is generated from logical expressions. The *logical specification*  $L$  consists of all formulas derived from a logical expression  $W_L$  using the algorithm  $\mathcal{A}$ , i.e.  $L(W_L) = \{f_i : i \geq 0 \wedge f_i \in \mathcal{A}(W_L, P)\}$ , where  $f_i$  is a TL formula. Generating logical specifications is not a simple summation of formula collections resulting from a logical expression. The generation algorithm has two inputs. The first one is a logical expression  $W_L$  which is a kind of variable, i.e. it varies for every (workflow) model, when the workflow is subjected to any modification. The second one is a predefined set  $P$  which is a kind of constant, i.e. once defined then widely used. The example of such a set is shown in Fig 8. All formulas describe both safety and liveness properties for a pattern [2]. However, the formulas are not discussed in the work



because they might be a subject of consideration in a separate work. Moreover, the formulas can and should be prepared by an expert with skills and theoretical background. It guarantees that an inexperienced software analyst or engineer will be able to obtain correct logical models. Most elements of the predefined  $P$  set, i.e. comments, two temporal logic operators, classical logic operators, are not in doubt. The slash allows to place more formulas in a single line.  $f_1, f_2$  etc. are atomic formulas for a pattern. They constitute a kind of formal arguments for a pattern.  $\Diamond f$  means that sometime (or eventually in the future), activity  $f$  is satisfied, i.e. the token reaches the activity.  $c(f)$  means that a logical condition associated with activity  $f$  has been evaluated and is satisfied. For example, the expression  $f \wedge c(f)$  means that the  $f$  activity is satisfied (executed) and  $c(f)$  is true (neither false nor undetermined). When the rising edge for the  $f$  activity is reached, the  $c(f)$  becomes undetermined until it is evaluated to true or false. When  $c(f)$  is undetermined, then the evaluation of the entire sample expression  $f \wedge c(f)$  is stopped until the evaluation is possible, i.e. all elements of the expression are determined.

The notion of joined entry and exit formulas is introduced which is a result of nested workflows, as well as the need to transfer, informally speaking, the logical signal to all start/termination points of a nested workflow. Let  $w$  is a logical expression, then  $w^\blacktriangleright$  is the *joined entry formula*, and  $w^\blacktriangleleft$  is the *joined exit formula*, when the joined formula is calculated using the following (recursive) rules:

1. if there is no workflow itself in the place of the first atomic formula/argument (the  $f_1$  formula), or the last atomic formula/argument (the  $f_4$  formula), respectively, then  $w^\blacktriangleright$  is equal to  $f_1$ , or  $w^\blacktriangleleft$  is equal to  $f_4$ , respectively,
2. if there is a workflow, say  $t()$ , in a place of the first argument, i.e. the  $f_1$  formula, or the last argument, i.e. the  $f_4$  formula, respectively, then it is replaced by  $t()^\blacktriangleright$ , or  $t()^\blacktriangleleft$ , respectively for every such case.

These rules allow to define joined point formulas for an arbitrary complex logical expression. For example, for logical expression  $w = Seq(a, b)$  formulas are  $w^\blacktriangleright = a$  and  $w^\blacktriangleleft = b$ . For expression  $w = Branch(a, b, c, Seq(d, e))$  formulas are  $w^\blacktriangleright = a$  and  $w^\blacktriangleleft = Seq(d, e)^\blacktriangleleft = e$  (by the way, for nested  $Seq(d, e)^\blacktriangleright = d$ ). For expression  $w = Seq(Concur(Seq(a, b), c, d, e), f)$  formulas are  $w^\blacktriangleright = Concur(Seq(a, b), c, d, e)^\blacktriangleright = Seq(a, b)^\blacktriangleright = a$  and  $w^\blacktriangleleft = f$ .

The output of the generation algorithm is a logical specification understood as a set of temporal logic formulas. The generation algorithm ( $\mathcal{A}$ ) is given as Algorithm 1. Let us assume some auxiliary constraints for the Algorithm. It is mandatory for every two patterns to have disjoint sets of atomic activities (arguments). Every pattern consists of at least two activities/tasks (arguments), c.f. Fig. 6 or Fig. 8. Let us supplement the Algorithm by some examples. The example for lines 3–5:  $Concur(a, b, c, d)$  gives  $L = \{a \Rightarrow \Diamond b \wedge \Diamond c, \neg a \Rightarrow \neg(\Diamond b \wedge \Diamond c), b \wedge c \Rightarrow \Diamond d, \neg(b \wedge c) \Rightarrow \neg\Diamond d, \Box\neg(a \wedge (b \vee c)), \Box\neg((b \vee c) \wedge d), \Box\neg(a \wedge d)\}$ . The example for lines 6–12:  $Branch(Seq(a, b), c, d, e)$  leads to  $L = \left[ \begin{array}{l} (a \vee b) \Rightarrow (\Diamond c \wedge \neg\Diamond d) \vee (\neg\Diamond c \wedge \Diamond d), \neg(a \vee b) \Rightarrow \neg((\Diamond c \wedge \neg\Diamond d) \vee (\neg\Diamond c \wedge \Diamond d)), c \vee d \Rightarrow \Diamond e, \neg(c \vee d) \Rightarrow \end{array} \right.$

---

**Algorithm 1** Generating logical specifications ( $\mathcal{A}$ )

---

**Input:** Logical expression  $W_L$  (non-empty), predefined set  $P$  (non-empty)**Output:** Logical specification  $L$ 

```
1:  $L := \emptyset$  ▷ initiating specification
2: for every workflow  $wrf()$  of  $W_L$  from left to right do
3:   if all arguments of  $wrf()$  are atomic then
4:      $L := L \cup wrf()$ 
5:   end if
6:   if any argument of  $wrf()$  is a workflow pattern itself then
7:     for every such an argument, say  $r()$ , substitute
8:       disjunction of its joined entry and exit
9:       formulas in all places where the argument
10:      occurs in the  $wrf()$  temporal formulas, i.e.
11:       $L := L \cup (wrf() \leftarrow "r() \blacktriangleright \vee r() \blacktriangleleft")$ 
12:   end if
13: end for
```

---

$\neg\Diamond e, \Box\neg((a \vee b) \wedge e), \Box\neg(c \wedge d), \Box\neg((a \vee b) \wedge (c \vee d)), \Box\neg((c \vee d) \wedge e)\} \cup \{a \Rightarrow \Diamond b, \neg a \Rightarrow \neg\Diamond b, \Box\neg(a \wedge b)\} \cup \{(a \vee b) \Rightarrow (\Diamond c \wedge \neg\Diamond d) \vee (\neg\Diamond c \wedge \Diamond d), \neg(a \vee b) \Rightarrow \neg((\Diamond c \wedge \neg\Diamond d) \vee (\neg\Diamond c \wedge \Diamond d)), c \vee d \Rightarrow \Diamond e, \neg(c \vee d) \Rightarrow \neg\Diamond e, \Box\neg((a \vee b) \wedge e), \Box\neg(c \wedge d), \Box\neg((a \vee b) \wedge (c \vee d)), \Box\neg((c \vee d) \wedge e), a \Rightarrow \Diamond b, \neg a \Rightarrow \neg\Diamond b, \Box\neg(a \wedge b)\}$ . The first set follows directly from lines [6–12], the second set follows directly from lines [3–5], while the [final] specification is the sum of all generated sets. Other examples are shown in Section 7.

**Remarks and discussion.** Let us supplement the approach and the Algorithm by some considerations. The evidence from the approach is that atomic formulas  $f_1$  and  $f_4$  play an important role for every pattern. They are always the first and the last, respectively, active activity/task for a pattern. It means they constitute the entry and the exit, respectively, point for a pattern. They are the first and the last, respectively, argument for a pattern. In the work, it is established that the maximum number of arguments for a pattern is  $n = 4$ , i.e. argument/formula  $f_4$ , c.f. Fig. 6 and Fig. 8. However, the limitation could be changed through introducing new predefined patterns which require, if necessary, greater number activities for a pattern. Predefining new patterns, one can introduce new types of iterations, more complex forks or branchings, and sequences with a greater number of activities comparing the patterns in Fig. 6. These new patterns must have only one entry and only one exit argument. Afterwards, the new patterns must be defined in terms of temporal logic formulas, c.f. Fig. 8.

The logical expression, defined in Section 6, is similar to the well-known regular expression. The internal structure of nested patterns (parentheses) is formed in such a way that for any two patterns, considered as arguments of the outer pattern, it is always satisfied that either the first and the second patterns are completely disjointed, or the first pattern is completely contained in the

second one, or the second patterns is completely contained in the first one. It follows from the recursive nature of the logical expression definition and correctly paired parenthesis.

The computational complexity of Algorithm 1 follows from the main loop which starts in line 2. The number of patterns processed from left to right in the whole logical expression is expressed as a linear function. If the considered pattern contains not only atomic arguments, then it is necessary to calculate both joined entry and exit formulas. On the other hand, if non-atomic arguments do not take the most outer positions, i.e.  $f1$  or  $f4$ , in the pattern, that it is not necessary to go outside the pattern, and it is enough to indicate the first or the last argument as joined entry or exit, respectively, formula. Otherwise, the search for joined formulas follows from the current position to the left or right, respectively, and in the worst-case to the beginning and to the end of the entire logical expression. On the other hand, such worst-cases should be rare. Summing up, computational complexity is linearly dependent on the number of patterns in the logical expression, and, in the worst-case, linearly from the calculation of joined formulas. Thus, the worst-case complexity is bounded above by quadratic function. For the average-case, the time complexity tends to the linear function.

A set of logical formulas is consistent if it does not contain contradiction, i.e. it does not contain any two provable formulas such that the first formula is a negation of the second formula. As it has already been said, logical patterns are predefined by a logician for further usage by an ordinary analyst. This assumption is made and the logician is responsible for consistency of predefined specifications. However, the open question is whether the Algorithm preserves consistency when generating logical specifications. In the case of lines 3–5, due to the disjointedness of atomic formulas for patterns, and the assumption of consistency of predefined patterns, it is not possible to introduce contradictions when adding a new elementary set of formulas. Let us note, that the following general formulas are valid and consistent:  $f1 \Rightarrow \Diamond f4$ , and  $\Box \neg(f1 \wedge f4)$ , which means that if  $f1$  (entry) is satisfied, then sometime in the future  $f4$  (exit) is satisfied, and that is always that  $f1$  and  $f4$  are not satisfied in the same time. Due to the consistency of the above general formulas and formulas for every pattern, the newly-generated logical specification is consistent, and it follows from the fact that new temporal formulas for joined entry and exit formulas refer to the consistent general/transition formulas of a pattern.

Completeness means that if a formula is true, it can be proven. However, as it has already been said in the case of consistency, the assumption is done that predefined specifications preserve completeness. The question is whether the Algorithm preserves completeness when generating logical specification. Let us note that all patterns in a logical expression are entirely nested, i.e. it is not possible to obtain a partial nesting that might provides an undesirable crossing of patterns. Moreover, every two pattern contain disjoint sets of atomic formulas. Completeness refers to the reachability all formulas and properties for every used logical pattern. The entry and exit formulas are generalization for a nested pattern and allow to bypass/skip its internal behaviour, if necessary, in other

words, they guarantee access to a pattern both to/from the “front” and to/from the “back” of a pattern with respect to both the preceding and the following pattern. On the other hand, considering the disjunction of entry and exit formulas for a pattern it allows to obtain/reach its entry end exit points, and consequently the remaining formulas. It may cause some redundancy of generated formulas, but on the other hand it covers all properties of combined patterns. Thus, the Algorithm does not introduce itself incompleteness to the output logical specifications with respect to predefined input specifications.

## 7 Reasoning and verification

Let us summarize the entire method proposed in the work, referring to the diagram at the end of Section 2. The first phase, let us call it the *modeling phase*, enables development of requirements models and includes the following steps:

- modeling of all use case diagrams  $UCD_1, \dots, UCD_m$ , where  $UC_1, \dots, UC_n$  are all use cases contained in all use case diagrams;
- modeling of scenarios for all use cases  $UC_1, \dots, UC_n$  and identification of atomic activities  $AA = \{a_1, \dots, a_l\}$ ;
- modeling of activity diagrams  $AD_1, \dots, AD_n$  for all scenarios using predefined workflow patterns, c.f. Fig. 6, and using the identified atomic activities.

All the above steps require the assistance of an engineer and cannot be done automatically. The next phase, let us call it the *analytical phase*, introduces a certain degree of automation and includes the following steps:

- translation of all activity diagrams  $AD_1, \dots, AD_n$  (and their workflows) to logical expressions  $W_{L,1}, \dots, W_{L,n}$ ;
- generation of logical specifications  $L_1, \dots, L_n$  for all logical expressions using the  $\mathcal{A}$  algorithm, i.e.  $\mathcal{A}(P, W_{L,i}) \rightarrow L_i$  for every  $i = 1, \dots, n$ ;
- summing of specifications, i.e.  $L = L_1 \cup \dots \cup L_n$ ;
- (manual) definition of the desired property  $Q$ ;
- start of the process of automatic reasoning using the semantic tableaux method for formula  $f_1 \wedge \dots \wedge f_k \Rightarrow Q$ , where  $f_1, \dots, f_k$  are formulas which belong to the logical specification  $L$ .

The above steps illustrate the entire operation of the system shown in Fig. 5. The loop between the last two steps, c.f. Fig. 1, refers to a process of both introducing and verifying more and more new properties (formula  $Q$ ) of the examined model.

Let us consider the activity diagrams  $AD_2$  and  $AD_3$  shown in Fig. 7 for use cases  $UC_2$  “InsuranceDamageLiquidation” and  $UC_3$  “VehicleCarriage”. Activity diagrams constitute the input for the deduction system shown in Fig. 5. The logical expression  $W_{L,2}$  for  $AD_2$  is

```
Seq(SystemLogIn, Branch(DamageVindication, Concur(
Seq(SupplyDocumentaryEvidence, InitRepair),
Seq(MechanicalRepairs, BodyRepairs), RentVehicle, TestDrive),
WarningDoubleVindication, SystemLogOut))
```

The logical expression  $W_{L,3}$  for  $AD_3$  is

$$Seq(Seq(CallAssistance, TransportVehicle), InitRepair)$$

Substituting letters of the Latin alphabet in places of propositions:  $a$  – Application,  $b$  – DamageVindication,  $c$  – SupplyDocumentaryEvidence,  $d$  – InitRepair,  $e$  – MechanicalRepairs,  $f$  – BodyRepairs,  $g$  – RentVehicle,  $h$  – TestDrive,  $i$  – WarningDoubleVindication,  $j$  – FinalReport,  $k$  – CallAssistance, and  $l$  – TransportVehicle, then the expression  $W_{L,2}$  is

$$Seq(a, Branch(b, Concur(Seq(c, d), Seq(e, f), g, h), i, j)) \quad (2)$$

and the expression  $W_{L,3}$  is

$$Seq(Seq(k, l), d) \quad (3)$$

Replacing propositions (atomic activities) by Latin letters is a technical matter and is suitable only for the work because of its limited size. In the real world, c.f. the deduction system from Fig. 5, original names of the activities would be used.

A logical specification  $L_3$  for the logical expression  $W_{L,3}$  is built in some steps that result from Algorithm 1. At the beginning, the specification of a model is  $L_3 = \emptyset$ . Patterns are processed from left to right. Hence, the first processed pattern is  $Seq$  and the next considered are:  $Branch$ ,  $Concur$ ,  $Seq$ , and  $Seq$ . The resulting logical specifications contain the formulas

$$\begin{aligned} L_2 = \{ & a \Rightarrow \Diamond(b \vee j), \neg a \Rightarrow \neg \Diamond(b \vee j), \Box \neg(a \wedge (b \vee j)), \\ & b \Rightarrow (\Diamond(c \vee h) \wedge \neg \Diamond i) \vee (\neg \Diamond(c \vee h) \wedge \Diamond i), \\ & \neg b \Rightarrow \neg((\Diamond(c \vee h) \wedge \neg \Diamond i) \vee (\neg \Diamond(c \vee h) \wedge \Diamond i)), \\ & (c \vee h) \vee i \Rightarrow \Diamond j, \neg((c \vee h) \vee i) \Rightarrow \neg \Diamond j, \Box \neg(b \wedge j), \Box \neg((c \vee h) \wedge i), \\ & \Box \neg(b \wedge ((c \vee h) \vee i)), \Box \neg(((c \vee h) \vee i) \wedge j), (c \vee d) \Rightarrow \Diamond(e \vee f) \wedge \Diamond g, \\ & \neg(c \vee d) \Rightarrow \neg(\Diamond(e \vee f) \wedge \Diamond g), (e \vee f) \wedge g \Rightarrow \Diamond h, \neg((e \vee f) \wedge g) \Rightarrow \neg \Diamond h, \\ & \Box \neg((c \vee d) \wedge ((e \vee f) \vee g)), \Box \neg(((e \vee f) \vee g) \wedge h), \Box \neg((c \vee d) \wedge h), \\ & c \Rightarrow \Diamond d, \neg c \Rightarrow \neg \Diamond d, \Box \neg(c \wedge d), e \Rightarrow \Diamond f, \neg e \Rightarrow \neg \Diamond f, \Box \neg(e \wedge f) \} \quad (4) \end{aligned}$$

and

$$\begin{aligned} L_3 = \{ & (k \vee l) \Rightarrow \Diamond d, \neg(k \vee l) \Rightarrow \neg \Diamond d, \Box \neg((k \vee l) \wedge d), \\ & k \Rightarrow \Diamond l, \neg k \Rightarrow \neg \Diamond l, \Box \neg(k \wedge l) \} \quad (5) \end{aligned}$$

Formulas 4 and 5 represent the output, i.e.  $L_2 \cup L_3 = L$ , of the  $\boxed{G}$  component in Fig. 5.

Formal *verification* is the act of proving the correctness of a system (liveness, safety). *Liveness* means that the computational process achieves its goals, i.e. something good eventually happens. *Safety* means that the computational

process avoids undesirable situations, i.e. something bad never happens. The liveness property for the model can be

$$b \Rightarrow \Diamond g \quad (6)$$

which informally/verbally means that **if the damage vindication is satisfied then sometime in the future the replacement car is reached**, formally  $DamageVindication \Rightarrow \Diamond RentVehicle$ . Another example, including extended use case, of the liveness property is

$$k \Rightarrow \Diamond h \quad (7)$$

which means that **if the call assistance is satisfied then sometime in the future the test drive is reached**, formally  $CallAssistance \Rightarrow \Diamond TestDrive$ . The safety property for the examined model can be

$$\Box \neg (i \wedge g) \quad (8)$$

which means that **it never occurs that the rental of a vehicle and the double vindication are satisfied in the same time**, or more formally  $\Box \neg (WarningDoubleVindication \wedge RentVehicle)$ . Another example of the safety property is

$$\Box \neg (i \wedge l) \quad (9)$$

which means that **it never occurs that the transport of a vehicle and the double vindication are satisfied in the same time**, or more formally  $\Box \neg (WarningDoubleVindication \wedge TransportVehicle)$ . When considering the property expressed by Formula 6, then the whole formula to be analyzed using semantic tableaux, providing a combined input for the  $\boxed{T}$  component in Fig. 5, is

$$(a \Rightarrow \Diamond(b \vee j)) \wedge \dots \wedge \Box \neg (k \wedge l) \Rightarrow (b \Rightarrow \Diamond g) \quad (10)$$

where the antecedent of the (main) implication is the conjunction of formulas from the generated sets  $L_2$  and  $L_3$  (Formulas 4 and 5), i.e. conjunction  $C(L) = C(L_2 \cup L_3)$ . In a similar way, the whole formula for Formula 7 is expressed by

$$C(L_2 \cup L_3) \Rightarrow (k \Rightarrow \Diamond h) \quad (11)$$

When considering the property expressed by Formulas 8 and 9, then the whole formulas are constructed in a similar way as

$$C(L_2 \cup L_3) \Rightarrow \Box \neg (i \wedge g) \quad (12)$$

and

$$C(L_2 \cup L_3) \Rightarrow \Box \neg (i \wedge l) \quad (13)$$

In all cases, i.e. Formulas 10, 11, 12, and 13, after the negation of the input formula within the prover, the inference trees are built. Presentation of a full inference tree for both cases exceeds the size of the work. (The simple inference tree from Fig. 4 gives an idea how it works.) All branches of the semantic trees are closed, i.e. Formulas from 6 to 9 are satisfied in the considered requirements model. In the case of falsification of the semantic tree the open branches are obtained and provide information about the source of an error what is another advantage of the method.

Although the logical specification was generated for only activity diagrams  $AD_2$  and  $AD_3$ , that is  $L = L_2 \cup L_3$ , c.f. Formula 4 and 5, the method is easy to scale up, i.e. extending and summing up logical specifications for other activity diagrams and their scenarios. Then, it will be possible to examine logical relationships (liveness, safety) for different activities coming from different activity diagrams for different use cases and scenarios.

## 8 Conclusions

The work proposes a two-phase strategy for formal analysis of requirements models. The first one is carried out by an engineer using a defined methodology and the second one can be (in most) automatic and enables formal verification of the desired properties (liveness, safety) of behaviour. The method for an automatic generation of logical specifications is proposed. This specification is a set of temporal logic formulas and obtaining it is crucial in the case of a practical use of the deduction-based formal verification.

The proposed method enables the construction in a formal way requirements models and then extracting logical specifications. The proposed method of generating enables scaling up, when building more and more nesting patterns. Logical patterns are once defined (logician) and could be commonly used by users (engineers). Introducing logical patterns as logical primitives allows for breaking of some barriers and obstacles in receiving logical specifications as a set of a large number of temporal logic formulas in an automated way. Application of formal verification, which is based on deductive inference, helps to significantly increase the maturity of requirements models considering infinite computations and using a human-intuitive approach.

Future works may include the implementation of the logical specification generation module. Another important issue could be a detailed analysis of the available provers [36] which could be useful and applied for the approach. It should result in a CASE software, e.g. Integrated Development Environments (IDEs), which could be a first step involved in creating industrial-proof tools, i.e. implementing another part of formal methods, hope promising, in industrial practice. The literature review argues that there is a lack of such comprehensive tools.

## References

1. Abrial, J.R.: Formal methods : Theory becoming practice. *Journal of Universal Computer Science* pp. 619–628 (2007)
2. Alpern, B., Schneider, F.B.: Defining liveness. *Information Processing Letters* 21 (4), 181–185 (1985)
3. Barrett, S., Sinnig, D., Chalin, P., Butler, G.: Merging of use case models: Semantic foundations. In: 3rd IEEE International Symposium on Theoretical Aspects of Software Engineering (TASE’09). pp. 182–189 (2009)
4. van Benthem, J.: *Handbook of Logic in Artificial Intelligence and Logic Programming*, chap. Temporal Logic, pp. 241–350. 4, Clarendon Press (1993–95)
5. Blanc, X., Mounier, I., Mougnot, A., Mens, T.: Detecting model inconsistency through operation-based model construction. In: 30th International Conference on Software Engineering (ICSE 2008), Leipzig, Germany, May 10-18, 2008. pp. 511–520. ACM (2008)
6. Booch, G., Rumbaugh, J., Jacobson, I.: *The Unified Modeling Language Reference Manual*. Addison Wesley (1999)
7. Cabral, G., Sampaio, A.: Automated formal specification generation and refinement from requirement documents. *Journal of the Brazilian Computer Society* 14 (1), 87–106 (2008)
8. Cao, L., Ramesh, B.: Agile requirements engineering practices: An empirical study. *IEEE Software* 25(1), 60–67 (2008)
9. Chakraborty, S., Sarker, S., Sarker, S.: An exploration into the process of requirements elicitation: A grounded approach. *Journal of the Association for Information Systems* 11(4), 212–249 (2010)
10. Chellas, B.F.: *Modal Logic*. Cambridge University Press (1980)
11. Chomicki, J., Saake, G. (eds.): *Logics for Databases and Information Systems*. Kluwer (1998)
12. Clarke, E., Grumberg, O., Peled, D.: *Model Checking*. MIT Press (1999)
13. Clarke, E., Wing, J., et al.: Formal methods: State of the art and future directions. *ACM Computing Surveys* 28 (4), 626–643 (1996)
14. d’Agostino, M., Gabbay, D., Hähnle, R., Posegga, J.: *Handbook of Tableau Methods*. Kluwer Academic Publishers (1999)
15. Dixon, C., Konev, B., Schmidt, R.A., Tishkovsky, D.: Labelled tableaux for temporal logic with cardinality constraints. In: 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2012), Timisoara, Romania, September 26–29, 2012. pp. 111–118 (2012)
16. Emerson, E.: *Handbook of Theoretical Computer Science*, vol. B, chap. Temporal and Modal Logic, pp. 995–1072. Elsevier, MIT Press (1990)
17. Eshuis, R., Wieringa, R.: Tool support for verifying uml activity diagrams. *IEEE Transactions on Software Engineering* 30 (7), 437–447 (2004)
18. Galton, A.: Logic as a formal method. *The Computer Journal* 35, 431–440 (1992)
19. Goré, R., Thomson, J., Widmann, F.: An experimental comparison of theorem provers for ctl. In: Combi, C., Leucker, M., Wolter, F. (eds.) 18th International Symposium on Temporal Representation and Reasoning (TIME 2011) Lübeck , Germany, September 12–14, 2011. pp. 49–56. IEEE (2011)
20. Gries, D., Schneider, F.B.: *A Logical Approach to Discrete Math*. Springer (1993)
21. Hähnle, R.: *Tableau-based Theorem Proving*. ESSLLI Course (1998)
22. Hinchey, M., Jackson, M., Cousot, P., Cook, B., Bowen, J.P., Margaria, T.: Software engineering and formal methods. *Communications of the ACM* 51(9), 54–59 (2008), <http://doi.acm.org/10.1145/1378727.1378742>



23. Hoare, T., Misra, J.: Verified software: Theories, tools, experiments. vision of a grand challenge project. In: Meyer, B., Woodcock, J. (eds.) *Verified Software: Theories, Tools, Experiments*, First IFIP TC 2/WG 2.3 Conference, VSTTE 2005, Zurich, Switzerland, October 10–13, 2005, Revised Selected Papers and Discussions. *Lecture Notes in Computer Science*, vol. 4171, pp. 1–18 (2005)
24. Hurlbut, R.R.: A survey of approaches for describing and formalizing use cases. Tech. Rep. XPT-TR-97-03, Expertech, Ltd. (1997)
25. Hustadt, U., Schmidt, R.A.: An empirical analysis of modal theorem provers. *Journal of Applied Non-Classical Logics* 9(4), 479–522 (1999)
26. Islam, M.Z., Mashiyat, A.S., Khan, K.N., Karim, S.M.M.: A tableau based automated theorem prover using high performance computing. *Journal of Computers* 7(3), 597–607 (2012)
27. Kazhamiakin, R., Pistore, M., Roveri, M.: Formal verification of requirements using spin: A case study on web services. In: *Proceedings of 2nd International Conference on Software Engineering and Formal Methods (SEFM 2004)*, 28–30 September 2004, Beijing, China. pp. 406–415 (2004)
28. Klimek, R.: *Advanced Methods and Technologies for Agent and Multi-Agent Systems*, *Frontiers of Artificial Intelligence and Applications*, vol. 252, chap. A Deduction-based System for Formal Verification of Agent-ready Web Services, pp. 203–212. IOS Press (2013), <http://ebooks.iospress.nl/publication/32843>
29. Klimek, R.: From extraction of logical specifications to deduction-based formal verification of requirements models. In: Hierons, R.M., Merayo, M.G., Bravetti, M. (eds.) *Proceedings of 11th International Conference on Software Engineering and Formal Methods (SEFM 2013)*, 25–27 September 2013, Madrid, Spain. *Lecture Notes in Computer Science*, vol. 8137, pp. 61–75. Springer Verlag (2013), [http://dx.doi.org/10.1007/978-3-642-40561-7\\_5](http://dx.doi.org/10.1007/978-3-642-40561-7_5)
30. van Lamsweerde, A.: *Requirements Engineering - From System Goals to UML Models to Software Specifications*. John Wiley & Sons (2009)
31. Nikora, A.P., Balcom, G.: Automated identification of ltl patterns in natural language requirements. In: *ISSRE 2009, 20th International Symposium on Software Reliability Engineering*, Mysuru, Karnataka, India, 16–19 November 2009. pp. 185–194. IEEE Computer Society (2009)
32. Pender, T.: *UML Bible*. John Wiley & Sons (2003)
33. Pohl, K.: *Requirements Engineering: Fundamentals, Principles, and Techniques*. Springer Publishing Company (2010)
34. Rauf, R., Antkiewicz, M., Czarnecki, K.: Logical structure extraction from software requirements documents. In: *19th IEEE International Requirements Engineering Conference (RE 2011)*, Trento, Italy, 29 August – 2 September 2011. pp. 101–110. IEEE Computer Society (2011)
35. Rossi, C., Enciso, M., de Guzmán, I.P.: Formalization of uml state machines using temporal logic. *Software and System Modeling* 3(1), 31–54 (2004)
36. Schmidt, R.: Website: accessible theorem provers, <http://www.cs.man.ac.uk/~schmidt/tools/> (2014), accessed on 6-January-2014
37. Schneider, G., Winters, J.: *Applying use cases: a practical guide*. Addison-Wesley (2001)
38. Shankar, N.: Automated deduction for verification. *ACM Computing Surveys* 41(4), 20:1–20:56 (2009)
39. Smith, M.H., Havelund, K.: Requirements capture with rcats. In: *16th IEEE International Requirements Engineering Conference*, ( RE 2008), 8–12 September 2008, Barcelona, Spain. pp. 183–192. IEEE Computer Society (2008)

40. Sommerville, I., Kotonya, G.: Requirements Engineering: Processes and Techniques. John Wiley & Sons, Inc., New York, NY, USA (1998)
41. Winkler, S., Pilgrim, J.: A survey of traceability in requirements engineering and model-driven development. *Software and Systems Modeling* 9(4), 529–565 (2010), <http://dx.doi.org/10.1007/s10270-009-0145-0>
42. Wolper, P.: The tableau method for temporal logic: an overview. *Logique et Analyse* 28, 119–136 (1985)
43. Wolter, F., Wooldridge, M.: Temporal and dynamic logic. *Journal of Indian Council of Philosophical Research* XXVII(1), 249–276 (2011)
44. Woodcock, J., Larsen, P.G., Bicarregui, J., Fitzgerald, J.: Formal methods: Practice and experience. *ACM Computing Survey* 41(4), 19:1–19:36 (2009)
45. Yu, E., Giorgini, P., Maiden, N., Mylopoulos, J.: Social modeling for requirements engineering. MIT Press (2011)
46. Zhao, J., Duan, Z.: Verification of use case with petri nets in requirement analysis. In: *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA 2009): Part II*. pp. 29–42. Springer-Verlag (2009)